The British Society for the Philosophy of Science

Induction and Scientific Realism: Einstein versus van Fraassen Part One: How to Solve the Problem of Induction
Author(s): Nicholas Maxwell
Source: *The British Journal for the Philosophy of Science*, Vol. 44, No. 1 (Mar., 1993), pp. 61-79
Published by: Oxford University Press on behalf of The British Society for the Philosophy of Science
Stable URL: http://www.jstor.org/stable/687850
Accessed: 16/03/2011 19:37

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=oup.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Oxford University Press and *The British Society for the Philosophy of Science* are collaborating with JSTOR to digitize, preserve and extend access to *The British Journal for the Philosophy of Science.*

http://www.jstor.org

Brit. J. Phil. Sci. 44 (1993), 61–79 *Printed in Great Britain*

Induction and Scientific Realism: Einstein Versus van Fraassen Part One: How to Solve the Problem of Induction

NICHOLAS MAXWELL

1 INTRODUCTION

In this three-part paper, my concern is to expound and defend a conception of science, close to Einstein's, which I shall call *aim-oriented empiricism*. I shall argue that aim-oriented empiricism has the following virtues. (i) It solves the problem of induction; (ii) it provides decisive reasons for rejecting van Fraassen's brilliantly defended but intuitively implausible constructive empiricism (van Fraassen [1980, 1985]); (iii) it solves the problem of verisimilitude, the problem of explicating what it can *mean* to speak of scientific progress given that science advances from one false theory to another; (iv) it enables us to hold that appropriate scientific theories, even though false, can nevertheless legitimately be interpreted realistically, as providing us with genuine, even if only approximate, knowledge of unobservable physical entities; (v) it provides science with a rational, even though fallible and non-mechanical, method for the discovery of fundamental new theories in physics. And in connection with the last point, in the third part of the paper I show that Einstein made essential use of aim-oriented empiricism in scientific practice in developing special and general relativity. Aim-oriented empiricism is thus a methodological view of proven scientific fruitfulness! I conclude by considering to what extent Einstein came explicitly to advocate aim-oriented empiricism in his later years.[1]

[1] Elsewhere I have argued that aim-oriented empiricism has implications not just for philosophy of science and for science itself, but for all of academic inquiry. Somewhat as Popper generalized *falsificationism* to create *critical rationalism*, a general notion of rationality with implications for social thought and life, I have generalized aim-oriented empiricism (a distinct improvement over falsificationism) to create *aim-oriented rationalism* (an improvement over critical rationalism), a general notion of rationality with radical implications for inquiry as a whole, and for social and political life: see Maxwell [1976, 1980, 1986, 1991], and especially [1984].

In this first part of the paper, I expound aim-oriented empiricism, and show how it solves the problem of induction. As I have already expounded aim-oriented empiricism in a number of places during the last twenty years (see Maxwell [1972, 1974, 1976, 1977, 1979, 1980], and especially [1984], Ch. 9), I shall make my exposition here as brief as possible.

2 AIM-ORIENTED EMPIRICISM

According to aim-oriented empiricism, science must presuppose, even in the context of justification, that the universe is comprehensible in some way or other. A comprehensible universe is a universe throughout which there exists *something*, invariant and unchanging, at all times and places, which in some sense controls, determines or is responsible for, all that which varies and changes from place to place and from time to time, and in terms of which diversity and change can be explained and understood. There are a number of ways in which we can imagine the universe to be comprehensible in this sense (Maxwell [1984], pp. 218–21). One way, for example, would be to imagine that the all-pervasive, all-powerful *something* which renders the universe comprehensible is God. Modern physics, from the time of Kepler and Galileo down to the present, presupposes (implicitly or explicity) that the universe is comprehensible in the more specific sense that some kind of unified pattern of physical law, characterizable in principle by means of some coherent, unified piece of physically interpreted mathematics, runs through all phenomena. It is this unified pattern of physical law which theoretical physicists today—following the great example of Einstein—seek to capture and characterize in terms of the ultimate, unified, true 'theory of everything'—that uniquely true physical theory which brings together all forces and fundamental physical entities into one unified scheme and, in principle, predicts and explains all phenomena. According to this conception of the comprehensibility of the universe, that which exists can be divided up into two parts which may be called the *invariant*, U, and the *variable*, V. U is precisely the same everywhere, at all times and places, whereas V is in general different at different times and places. The crucial point, however, is that the way in which V varies in time is precisely specified by U—deterministically or probabilistically. If V changed differently, U could not exist. U thus determines how things change; change can be explained and understood once the nature of U is known. In order to specify U we need to discover the ultimately true, unified, comprehensive physical 'theory of everything'—let us call it T. The nature of U is specified by the character, the form, of T—that which is required and is the same in all applications of T. V is the variable physical state of any system or space–time region, specified in the language of T. Roughly, whereas U is what exists in Nature corresponding to the invariant form or character of T, V, at any given instant and spatial region, is the initial and boundary conditions, the (changeable) physical state, of what exists at the instant and region in question, specified in terms of T. If the universe consisted only of point-particles interacting by means of Newtonian gravitation, U would be the unchanging physical properties (inertial mass, gravitational charge), specified by the equations $F=ma$, $F=m_1m_2/r^2$, $F_{1+2}=F_1+F_2$, whereas V would be the instantaneous relative positions and velocities, and the masses, of the point-particles at the instant and spatial region in question.

Let us call the thesis that the universe is comprehensible in this more or less specific, physicalistic sense *physicalism*.[2]

The basic *aim* of theoretical physics, even in the context of justification, is to formulate and confirm physicalism as a precise, complete, true physical 'theory of everything'. That is, the aim is to specify the nature of U so that it is apparent, in principle at least, how U determines the way V changes in all physically possible circumstances. The basic *method* is to move progressively towards this goal by striving to formulate and confirm physical theories which—even though applicable only to a limited range of phenomena with only limited accuracy—nevertheless successively apply to wider and wider ranges of phenomena with ever increasing accuracy, thus capturing more and more of physicalistic reality, until the ultimate goal, T itself, is reached. This means that a theory, in order to be accepted as a part of scientific knowledge, must satisfy two requirements, namely (1) the requirement of sufficient *unity* (as it may be called) and (2) the requirement of sufficient empirical success. In order to satisfy (1), a theory, T_0, must be such that it promises to constitute a stepping stone towards T. It must be such that it can be interpreted as postulating the existence of something, U_0, which remains invariant throughout a range of apparently diverse phenomena, and which determines how things change in this range of phenomena. In order to satisfy (1), T_0 must promise to do this more accurately for a wider range of phenomena than any other already accepted theory. In order to satisfy this first requirement of *unity*, in short, T_0 must be internally unified, conceptually coherent, non-*ad hoc*, explanatory; T_0 must promise to help unify the whole of theoretical physics. In order to satisfy requirement (2), T_0 must predict phenomena sufficiently successfully, and in any case more successfully than any already accepted theory.

[2] What it means to assert that the universe is comprehensible in this physicalistic sense will be further clarified during the course of the paper. Physicalism, as understood here, does not imply that *everything that exists* is exclusively physicalistic in character, it being possible, in principle, to describe, explain and understand everything in physicalistic terms. As I have argued elsewhere (Maxwell [1966, 1968, 1976] and [1984], Ch. 10), we need to construe physics as being concerned only with a highly selective *aspect* of all that which exists—that aspect which determines how events unfold. Physics necessarily excludes all mention of the experiential dimensions of reality: colours, sounds and smells as we experience them, inner experiences, thoughts, states of consciousness.

At any given stage of its development (since the time of Kepler and Galileo at least) science has presupposed a more or less specific version of physicalism, which we may call the current 'metaphysical blueprint' of science. (It is a blueprint for the ultimate nature and comprehensibility of the universe.) Examples of metaphysical blueprints from the history of physics are: the corpuscular hypothesis of the seventeenth century; Boscovich's point-atomism; Einstein's unified field view; Wheeler's geometrodynamics. An example of a blueprint from contemporary physics is superstring theory, which is, at its present state of development, very much a blueprint rather than a theory. Some such scientifically accepted metaphysical blueprint is implicit in the current basic concepts of physics about such things as space, time, physical entity, force, mass, energy, momentum. The metaphysical blueprint that is accepted at any given stage exercises a profound influence over both what theories are accepted in the context of justification and what hypotheses are explored in the context of discovery. (The metaphysical blueprint determines the current non-empirical *methods* of science—methodological rules that include symmetry, invariance and conservation principles used both in formulating and in assessing new theories.) Almost certainly, however, the metaphysical blueprint accepted at any given stage is *false*—even if the general thesis of physicalism is true. It is thus essential, for the sake of scientific progress, that the currently adopted metaphysical blueprint be explicitly formulated so that it may be critically assessed, and so that modifications and alternatives may be considered, within science itself. This is required for scientific rigour—since quite generally an elementary requirement for intellectual rigour is that influential and problematic assumptions be made explicit, so that they can be critically assessed in the hope that they can be *improved*.

As the current metaphysical blueprint is modified, in the hope that it comes closer to representing the way in which the universe is comprehensible, so too the non-empirical methods or criteria employed in assessing the acceptability of theories (methods that include symmetry, invariance and conservation principles) are modified as well. (Thus in adopting the corpuscular metaphysical blueprint we adopt as well the methodological prescription that an acceptable theory is an action-by-contact theory; in rejecting this blueprint in favour of Boscovich's point-atom blueprint, we reject the corpuscular methodological rule in favour of the new rule: an acceptable theory is an action-at-a-distance theory.) Scientific rigour, in short, demands that the more or less specific aim and methods of science evolve with evolving knowledge. (Evolving blueprints leads to evolving *aims* for science for the simple reason that the basic *aim* of science, at any given stage, is to predict and explain phenomena in terms of the current *blueprint*: the aim is formulated in terms of the blueprint. Change the blueprint and we change the aim.) There is, in other words, a persistent interplay between best theories, best metaphysical blueprint or aim, and best methods or principles (such as symmetry, invariance and conservation principles) employed both in developing and in assessing theories. In improving the aim and methods of science we improve *knowledge about how to improve knowledge*. And it is this extraordinarily fruitful interplay between improving knowledge and improving knowledge about how to improve knowledge which, according to aim-oriented empiricism, is of the very essence of scientific method and rationality, and which has made possible the unprecedented explosive growth in scientific knowledge and understanding since the seventeenth century. (Improving knowledge leading to improving observational and experimental methods, in turn leading to the further improvement of knowledge, also plays a vital role in the explosive growth of science.) As long as the wrong idea as to how the universe is comprehensible is upheld (Aristotelianism, for example, or animism), there is little hope of rapid progress in scientific knowledge. Rapid progress begins to be achieved when the approximately correct view as to how the universe is comprehensible is adopted and pursued—physicalism or, as expressed by Galileo, the idea that the book of Nature 'is written in the language of mathematics' (simple mathematical laws governing motion and change). Proposing and testing empirically laws and theories which comply with the basic tenet of physicalism lead to ever accelerating progress in knowledge—this requiring, however, that metaphysical blueprints and associated methods or principles evolve with evolving knowledge within the general framework of physicalism.

Aim-oriented empiricism stands in stark contrast to *standard empiricism*—the doctrine that science makes no permanent presuppositions about the nature of the universe but proceeds in accordance with '*the principle of empiricism*, which asserts that in science, only observation and experiment may decide upon the *acceptance and rejection* of scientific statements, including laws and theories' (Popper [1963] p. 54). The conceptions of science that emerge from aim-oriented empiricism (AOE) and standard empiricism (SE) differ in the following respects. (1) AOE holds that science presupposes that the universe is comprehensible in some way or other and, since the seventeenth century, that it is comprehensible along the lines of physicalism; SE holds that science must make no such permanent metaphysical presupposition about the nature of the universe. (2) AOE holds that the intellectual integrity of science requires that the presupposition of comprehensibility be explicitly acknowledged within science; SE holds that the intellectual integrity of science requires that no such presupposition be made by science. (3) SE holds that theories be selected impartially with respect to empirical success or failure, any bias in favour of simplicity, unity, comprehensibility or some metaphysical paradigm or hard core (as described by Kuhn [1962] and Lakatos [1970]) being temporary, and not a permanent bias in favour of any thesis about the nature of the world; AOE holds that selection of theories is permanently biased in favour of the metaphysical thesis that the universe is comprehensible, indeed even (perhaps) permanently biased in favour of physicalism. (4) SE holds that scientific knowledge is made up of two domains, namely (i) empirical data and (ii) theory; AOE holds that scientific knowledge is made up of *five* domains, namely: (i) empirical data, (ii) theory, (iii) metaphysical blueprint, (iv) physicalism and (v) the general thesis that the universe is comprehensible in some way or other. (5) SE holds that science has a fixed aim, in the context of justification, and fixed methods for the selection of theories; AOE holds that science has evolving aims and methods in the context of justification. According to AOE, only at a sufficient level of generality does science have the *fixed* aim of improving knowledge of how the universe is comprehensible; only at the *metamethodological* level[3] does science have fixed *meta*methods which specify how metaphysical blueprints, methods and theories need to be modified in the light of each other (Maxwell [1974], pp. 257–64, [1984], pp. 236–7). (6) SE holds that the philosophy of science is not itself a part of science since it is not itself empirically testable; AOE holds that philosophies of science—views about what the more or less specific aim and methods of science ought to be—are a vital integral part of science, it being essential that there is interplay between theories, and aims-and-methods, in both directions. (7) According to AOE, the *rationality* of science cannot be divorced from its history (because of the interplay between theories and aims-and-methods); according to SE there is nothing especially historical in character about the rationality of science (science having a fixed aim and fixed methods). (8) AOE holds that there is a fallible, non-mechanical method of discovery within science even at the most fundamental level: metaphysical blueprints are modified and made more precise until they become testable (and empirically successful) theories; SE does not provide the means for any kind of rational method of discovery at the most fundamental level: according to SE, the most scientifically acceptable metaphysical ideas would be those compatible with existing theoretical knowledge, whereas fundamental new theories almost invariably emerge from ideas profoundly *incompatible* with pre-existing theories.

The scientific enterprise, as conceived of by aim-oriented empiricism, is depicted in Figure 1.

[3] The idea of characterizing scientific method on the 'metamethodological' level was first introduced by me in 1974 in my 'The Rationality of Scientific Discovery' paper. Thus in one place I remark: 'the reason why it is absolutely obligatory to characterize scientific method on what up until now has been thought of as the 'metamethodological' level . . . can be brought out quite decisively from the following elementary consideration . . . *different aims give rise to different rational rules of theory acceptance*' (Maxwell [1974], p. 260). The outer trappings of the idea, and some of the terminology, have been taken up subsequently, without acknowledgement, by L. Laudan (see, for example, Laudan [1990]). In Laudan's account, however, the central idea of committing science to the assumption that the universe is comprehensible is missing. There is thus no solution to the problem of induction, and the basic *raison d'être* for characterizing science in terms of progressively *improving* aims and methods disappears.

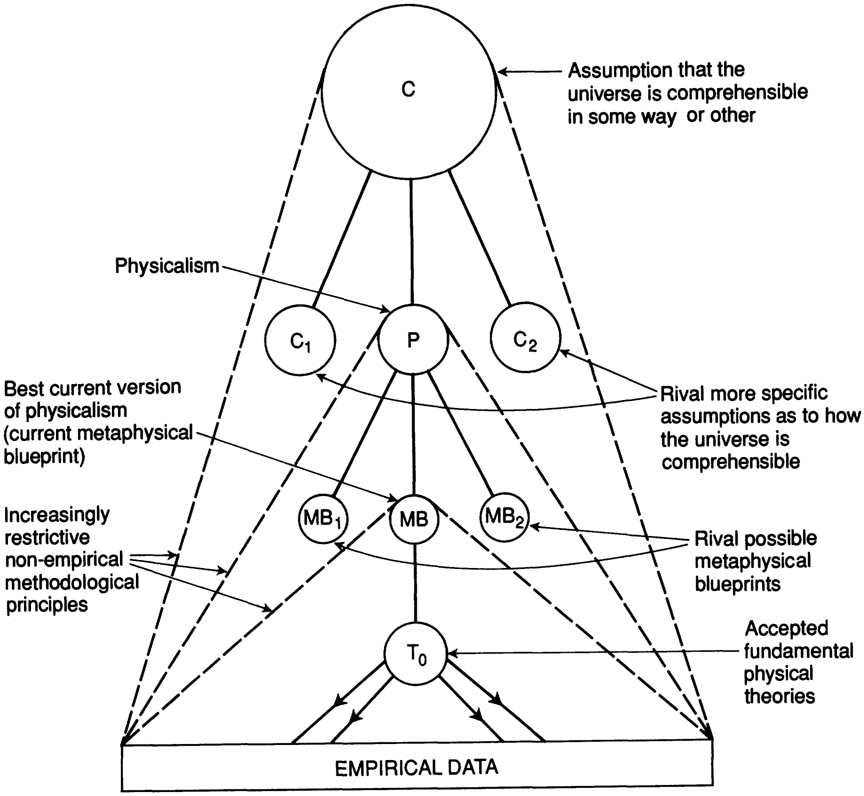

FIGURE 1. Aim-oriented empiricism.

3 AIM-ORIENTED EMPIRICISM REQUIRED TO SOLVE PROBLEM OF INDUCTION

All versions of standard empiricism fail disastrously to solve the problem of induction, and must therefore be rejected. In outline, the argument in support of this contention proceeds as follows.

Let T_0 be any physical theory sufficiently explanatory and empirically successful to be acceptable as a part of scientific knowledge. However empirically successful T_0 may be, there will always be infinitely many rival theories $T_1, \ldots T_\infty$, all at least as empirically successful as T_0 if not more so, any number of which can easily be formulated.

All we need to do in order to formulate such a theory is to proceed as follows. Let D be the domain of possible phenomena to which T_0 applies (the phenomena in principle predicted by T_0). Specify some subdomain d, within D, of possible phenomena *not yet observed*, in such a way that the predictions of T_0

fall either into d or into D–d, in a mutually exclusive way. The new rival 'aberrant' theory, T_1 say, consists of the following two postulates: (1) within D–d, everything is as T_0 predicts; (2) within d the phenomena are as P_1 predicts (and here P_1 may be whatever we please, but in any case different from what T_0 predicts in d). Suppose, for example, that T_0 is Newtonian theory. We can choose d to be some region of space and time containing gravitating bodies but not as yet observed, and P_1 might be $F = Gm_1m_2/r^3$, for bodies within d. Alternatively, we can choose d to be, not some region of space and time, but rather some region, within D, of *possible phenomena* which have not as yet been observed, or even become actual at any time or place. Thus d might consist of phenomena of the following type: two pure solid gold spheres, each of mass greater than a thousand tons, orbit around each other in outer space far away from other bodies. Here P_1 may be taken to be, as before, $F = Gm_1m_2/r^3$ for gold spheres in d. In this case in order to ensure that the abberant version of Newtonian theory has as great an empirical content as non-aberrant Newtonian theory, it will be necessary to specify d very narrowly and precisely, to ensure that there is no ambiguity as to whether any possible system falls within d or not.

As these examples make clear, however extensive the observational data may be in support of an accepted physical theory T_0, the range of possible phenomena predicted by T_0 which have not as yet been observed will always be infinite in extent. There will always be infinitely many non-overlapping subdomains, $d_1, d_2, \ldots d_\infty$, in D, of unobserved phenomena predicted by T_0; hence infinitely many aberrant theories $T_1 \ldots T_\infty$ exist which agree precisely with T_0 for all observed phenomena but disagree with T_0 for some unobserved phenomena. We can set out to refute such aberrant theories empirically—and if T_0 is a good theory, no doubt its aberrant variants will turn out to be false. There will always remain, however, infinitely many unrefuted aberrant rivals to T_0, all just as empirically successful as T_0 itself.

In practice, of course, empirically successful aberrant theories are dismissed out of hand as being too grotesquely *ad hoc* to be worthy of consideration. Such 'theories' are not even formulated in science; they fail to satisfy the kind of requirements a set of statements must satisfy if it is to constitute a candidate for an authentic scientific theory. The grounds for rejecting empirically successful aberrant theories are, in short, in practice, overwhelming and devastating, and any acceptable conception of science must be able to endorse and justify this. It is just this which standard empiricism fails lamentably to do. Any *honest* attempt to select theories *solely* on the basis of empirical success and failure, in an impartial way, no permanent presupposition being made about the nature of the universe, must give equal weight to the equally empirically successful theories T_0 and $T_1 \ldots T_\infty$. In persistently *accepting* theories like T_0 and *rejecting* theories like $T_1 \ldots T_\infty$, on grounds that have nothing to do with evidence, we thereby persistently presuppose that *the world itself* is such that aberrant phenomena, of the type postulated by aberrant theories, do not exist—however we may try to rationalize such a persistent presupposition away. The fact that in science infinitely many empirically successful aberrant theories are persistently and decisively dismissed without a scrap of evidence shows that standard empiricism fails disastrously as an account of scientific practice. The fact that empirically successful aberrant theories *must* be so dismissed if science is to progress shows that standard empiricism fails disastrously as a rational ideal for science. Science both does, and must, make the permanent metaphysical presupposition that the world is non-aberrant.

This argument becomes much more powerful when we take into account that in science even the most successful theories have their empirical limitations. Given any empirically successful theory T_0, we can distinguish four domains of phenomena: phenomena A, successfully predicted by T_0; phenomena B, falling within the scope of T_0, but not yet predicted (because here the equations of T_0 have not yet been solved for example); phenomena C, observationally established, that seem to clash with the predictions of T_0, and which thus ostensibly refute T_0; phenomena D, which fall outside the domain of applicability of T_0, and which are not predicted by any other accepted theory. Employing the methods indicated above, we can formulate a new theory, T_1, which asserts: in A and B, everything occurs as T_0 asserts; in C such and such phenomena occur (the observed phenomena); in D such and such phenomena occur (the observed phenomena). (A phenomenon is here taken to be a low level empirical law.) All the phenomena that support T_0 also support T_1; T_1 has greater empirical content than T_0; whereas T_0 is (ostensibly) refuted, T_1 is not; furthermore T_1 successfully predicts phenomena that refute T_0; and finally T_1 successfully predicts phenomena that lie beyond the scope of T_0. In these circumstances empirical considerations must overwhelmingly favour T_1 over T_0—and yet of course in practice, quite properly, T_1 would not even be formulated, let alone taken seriously as a possible rival to T_0. It is difficult to see how any version of standard empiricism can do justice to such persistent rejection of aberrant theories *in the teeth of the evidence.*

As an example of the general situation just described, consider Newtonian theory around 1890, after the discovery of the precession of the perihelion of the orbit of Mercury but before the development of special or general relativity. Newtonian theory is, on the face of it, refuted. An aberrant version of Newtonian theory can however be formulated which successfully predicts the observed precession of the perihelion of Mercury's orbit. This aberrant theory asserts that everything occurs in accordance with Newtonian theory except for systems that satisfy the following conditions: there are two bodies, one with mass and dimensions very close to that of the sun, the other moving in an orbit very close to that of Mercury, and with a mass no larger than that of Mercury. In this case the orbit of the second body precesses at precisely the rate observed for Mercury. Not only is this aberrant version of Newtonian theory empirically successful where Newtonian theory is falsified: in addition it predicts new phenomena, for sun-Mercury type systems, predictions which happen to be correct! Despite this empirical success, such a grotesquely *ad hoc* theory would not be taken seriously in science for a moment.

Ironically enough, in 1890 an *ad hoc* modification of Newtonian theory was put forward which was capable of predicting Mercury's precessing perihelion. This theory, put forward by Maurice Lévy, combined in an *ad hoc* way two distinct modifications of Newton's law of gravitation, one based on the way Weber had proposed Coulomb's law should be modified, the other based on the way Riemann had proposed Coulomb's law should be modified (see North [1965]). In discussing the theory, North comments: 'Needless to say, this approach to the problem was recognized for what it was worth, not least by its author' (North, p. 47). Even though it reproduced all the empirical success of Newtonian theory and was successful where Newtonian theory was refuted, and even though it was very much less *ad hoc* than the kind of aberrant theories we have been considering so far, nevertheless Lévy's theory was not taken seriously for a moment.

It is important to appreciate that the problem of providing a rationale for dismissing empirically successful aberrant theories (and the aberrant predictions of such theories) does not just arise in the context of pure scientific research, when the task is to choose the best theory from the standpoint of further research. The problem also arises in the context of practical, technological applications of science, when the task is to predict scientifically familiar phenomena as reliably as possible. However routine and standard such an application of a theory may be (in connection with bridge-building, for example), there is a sense in which the theory is always being applied to a domain d of as yet unobserved phenomena. The phenomena are unobserved in *both* senses: they occur at a time not yet observed; and they must, in what would ordinarily be regarded as inessential details, be different from all phenomena to which the theory has been previously applied—since no phenomenon, or state of affairs, is ever precisely repeated in all its details. (No two bridges are precisely alike, especially if we take their environments into account.) This means that whenever a well-established theory is used to predict phenomena as reliably as possible, in a practical, technological context, there will always exist rival aberrant theories, of both types, which will make dramatically different predictions for the phenomena in question. In practice we dismiss such aberrant theories and predictions out of hand, even though the theories may be just as empirically successful as, or even more successful than, the theory we actually accept and employ.

My claim is that no version of standard empiricism can do justice to, or provide a rationale for, the way in which in science and technology ostensibly refuted but explanatory theories are persistently chosen in preference to unrefuted and empirically more successful aberrant theories—the latter usually not even being formulated, or if formulated never being taken seriously.

We cannot simply argue, let it be noted, that aberrant theories are unacceptable because they fail to satisfy the requirment of strict universality (Popper [1959], pp. 62–70), according to which any acceptable law or theory must make no reference, explicit or implicit, to any special time, place or object. We cannot argue this for two reasons. In the first place, in demanding that acceptable theories satisfy this requirement we are in effect making a permanent metaphysical assumption about the world—one which is implicit in our methodology—according to which no special places, times or objects exist from the standpoint of lawful regularities. This is incompatible with standard empiricism. Furthermore, adopting this metaphysical assumption permanently and dogmatically is irrational. The assumption could be false. The world could be such that special times, places or objects do exist from the standpoint of lawful regularities. A science which excludes this possibility *a priori*, thus rendering it unknowable, cannot be rational. Indeed, by no means all important physical theories have satisfied the requirement. Kepler's laws of planetary motion, when first formulated, referred explicitly to a special object, the sun. In the second place, adopting the *a priori* requirement of strict universality suffices only to exclude aberrant theories of the first type; it fails to exclude those of the second type (for which d is specified in strictly universal terms). In short, the methodological rule of strict universality, upheld in a rigid, *a priori* fashion, is both too restrictive to be rational, and not restrictive enough to exclude aberrant theories of the second type.

One approach to the problem is that of classical instrumentalism. The persistent preference in science for simple, unifying or explanatory theories to complex, non-explanatory or aberrant theories is entirely harmless since at the level of theory there are no claims to scientific knowledge. Only at the observational level is there scientific knowledge: the task of theory is to organize or systematize in the best possible way this scientific knowledge of the observational. Naturally that theory is chosen which best organizes or systematizes empirical knowledge: persistent favouring of such 'explanatory' theories over 'non-explanatory' theories does not at all imply that science implicitly presupposes the world itself has some kind of comprehensible structure. (It is this line of thought, I suggest, which lies behind Pierre Duhem's *The Aim and Structure of Physical Theory*—a line of thought which van Fraassen has refurbished and defended with new arguments.)

This approach would no doubt succeed if the problem were merely to provide a rationale for choosing that theory best able to systematize a given body of empirical data, knowledge about observable but unobserved phenomena not being at issue. But this is not the problem we have been discussing at all. What is at issue is rather the following: confronted by two theories, T_0 an explanatory theory, and T_1 an aberrant theory, which are equally successful empirically, but which yield different predictions for some domain d of as yet unobserved phenomena, what is the rationale for deciding that the predictions of T_0 are correct and trustworthy, while those of T_1 deserve no consideration at all? What can be the rationale for deciding this when T_1 is actually more successful empirically? The above instrumentalist line of argument fails hopelessly to solve this problem.

Van Fraassen does not claim to solve the problem of induction. He does however claim that *explanatoriness*, in a sense which goes beyond mere empirical strength, even though a virtue in a theory, legitimately influencing the provisional acceptance of a theory (for the sake of further research for example), is very definitely not an *epistemic* virtue, legitimately influencing judgements of knowledge, truth or empirical adequacy (van Fraassen [1980], pp. 92–6). This doctrine, essential to van Fraassen's constructive empiricism, is of course no more than a rewording of 'classical instrumentalism' as I have called it, the position I have just indicated and refuted. Van Fraassen's doctrine must be rejected for precisely the same reason. Persistently and inevitably, in the context of technological applications as well as scientific research, we accept predictions of *explanatory* theories and reject aberrant predictions of aberrant theories, even when the aberrant theories have greater empirical strength and have met with greater empirical success. In the most blatant fashion imaginable, explanatoriness has massive epistemic import.

A second approach is that of Karl Popper [1959, 1963]. In science and technology we are rationally entitled to choose that theory which is the most falsifiable and has survived the severest testing. The most falsifiable theory is also, however, the most explanatory. Thus in giving preference to explanatory theories over non-explanatory, aberrant theories on ostensibly non-empirical grounds, we are merely giving preference to those theories that are the most falsifiable, that are the most vulnerable to empirical appraisal. Far from violating 'the principle of empiricism', we are proceeding in such a way as to maximize the capacity of experience to decide choice of theory in an impartial way; we are not in any way covertly presupposing that the world itself is non-aberrant.

This ingenious suggestion might be successful if greater falsifiability (or empirical content) always implied greater explanatoriness. But this is not the case. As our discussion has already shown, a highly explanatory theory may be given greater falsifiability, or empirical content, by the addition of independently testable *ad hoc* postulates; this drastically decreases—indeed utterly sabotages—the explanatoriness of the theory. As it happens, Popper came to recognize that a new theory, in order to be a candidate for consideration, must be simple or explanatory in a sense which does not reduce to mere falsifiability (Popper [1963], p. 241); what Popper has so far failed to recognize is that this concession destroys his proposed solution to the problem of induction (see Maxwell [1974, 1979] and [1984], Ch. 9).[4]

A third approach is to argue that whenever two theories, one simple or explanatory, the other complex, *ad hoc* or aberrant, seem to be supported equally by the same evidence, in reality (other things being equal) it is always the simple, explanatory theory which receives the greater empirical support (see Jeffreys [1957]; Barker [1957]; Harman [1965, 1968]; Lycan [1988]). What looks like persistent preference for non-aberrant, explanatory theories on non-empirical grounds is in reality persistent preference for theories that have received the best empirical support.

I find this line of argument entirely unconvincing. (Here van Fraassen and I agree.) I can see no reason why explanatory theories should be intrinsically more verifiable than non-explanatory theories, in the required way. If the phenomena under investigation are such that they are amenable to being explained, in the relevant sense of 'explanation' (whatever this may be), then to give preference to explanatory over non-explanatory theories when empirical support seems to be equal will meet with success. But if the phenomena are *not* explainable in this sense, this policy will meet with persistent failure. In short, persistently 'to infer to the best explanation' is just to presuppose, on non-empirical grounds, that the phenomena themselves are comprehensible or explainable.[5]

It might be argued that given a theory T has been verified for some phenomena in D–d but not in d, then this only provides empirical support for the predictions of T within d if the *form* of T is the same throughout D–d and d. This holds for any non-aberrant explanatory theory T_0 but not for aberrant theories of the kind indicated above. We are thus entitled to ignore aberrant predictions of empirically successful aberrant theories.

This suggestion does not solve the problem. It is, in the first place, not at all obvious what 'invariant form' is supposed to mean here. Granted that we have before us two theories, the first T_0, an explanatory theory with an invariant form throughout its domain of application, the other T_1, an aberrant theory

[4] Watkins' attempt to rescue Popper from his critics (see Watkins [1984] does not help. Watkins does, it is true, put forward a requirement of unity or 'organicity' that acceptable scientific theories must satisfy. This demands, roughly, that an acceptable scientific theory must have greater empirical content than the total empirical content of any two parts of the theory, given that the theory has been 'naturally' axiomatized (Watkins [1984], section 5.3). Unfortunately, empirically successful but severely aberrant theories, of the type considered in the present paper and in earlier publications of mine, are perfectly capable of satisfying this requirement. It is rather striking that, despite the sustained attention that Watkins devotes to questions about the *aims* of science, he fails to take into account, or even note, the highly relevant work on this subject that I published a decade earlier (Maxwell [1974]).

[5] For a somewhat more detailed criticism of 'induction to the best explanation', see Maxwell (forthcoming).

with an abruptly changing form within d, we could always reformulate the theories in such a way that it is T_0 that abruptly changes its form within d, and T_1 that has an invariant form throughout D. (This generalizes Goodman's considerations concerning 'grue' and 'bleen': Goodman [1954].) The decision to stick to our customary concepts so as to exclude this relativity of aberrance, and at the same time dismiss empirically successful aberrant theories from consideration, in the way indicated, amounts to presupposing permanently that the phenomena themselves are 'non-aberrant'. But even if we do this, we still face the problem of specifying precisely what is to count as an 'aberrant' theory. It is clearly not sufficient to demand that the functional relationships that constitute the theory are *continuous* throughout D, since aberrant theories of the type considered above could easily be reformulated so as to satisfy this condition. We might try demanding that the functional relationships of T are *analytic*. The idea here is this. Granted that we have determined precisely the functional relationships of T within any small region r within D, and granted that these functional relationships are analytic, then this fixes, by analytic continuation, the form of T for the whole of D, including d. An acceptable theory, in other words, must be such that, once it has been determined precisely for any small region within its domain of application, its form is determined for all phenomena to which it is applicable. Unfortunately this proposal suffers from two defects. In the first place, requiring that all acceptable physical theories must be analytic clearly involves making a restrictive assumption about the nature of the phenomena under investigation on non-empirical grounds. It involves the repudiation of standard empiricism. But secondly, this *a priori* presupposition is not restrictive enough to exclude theories empirically indistinguishable from aberrant theories. Given an analytic theory T_0, and given an aberrant theory T_1 which agrees precisely with T_0 throughout D–d but differs from T_0 throughout d (but which is continuous, we may suppose, within d), then there exists an analytic theory T_2, which differs only infinitesimally from T_0 throughout D–d and only infinitesimally from T_1 within a region d* within d. In other words, analytic theories always exist which imitate aberrant theories as accurately as we please, and certainly within experimental error.[6] The more we try to solve this second problem by placing more and more restrictive *a priori* constraints on theories, the more we run afoul of the first problem—and the further and further we depart from standard empiricism.[7]

[6] I am appealing here to the Stone-Weierstrass theorem which tells us, in effect, that any continuous function can be approximated arbitrarily closely throughout a finite interval by analytic functions: see Dieudonné [1960], pp. 131–4. I am grateful to Robert Seymour for drawing my attention to this theorem and reference.

[7] Other approaches, such as those of bootstrapping (Glymour [1980]), Bayesianism (Howson and Urbach [1989]), and anti-deductivism (Stove [1982]) do not help at all.

What we need to do, in order to overcome this impasse, is to appreciate that all forms of standard empiricism are untenable. They exclude the possibility of solving the problem of induction. In persistently preferring non-aberrant, explanatory theories to empirically more successful aberrant theories, in science and technology, we persistently bias the choice of our theories to accord with the metaphysical presupposition that the world itself is non-aberrant, or comprehensible. Indeed, inevitably at any given stage in the development of science, we bias the choice of theories to accord with a metaphysical assumption about the nature of the world—the current 'metaphysical blueprint'—which is very much more *specific* than the general assumption of non-aberrance or comprehensibility, and which is almost certainly *false*. In order to proceed in an intellectually rigorous way—in a way which gives us the best chance of improving scientific knowledge—it is essential that we make the current best metaphysical blueprint *explicit*, so that it can be critically assessed in the light both of empirical progress and the general assumption of comprehensibility, in the hope that it can in this way be *improved*. (Quite generally, intellectual rigour requires, at the very least, that we make explicit and so criticizable assumptions that are substantial, influential, problematic and implicit.) This interplay between improving empirical and theoretical knowledge on the one hand, and improving metaphysical blueprint and non-empirical methods (principles or criteria of explanatoriness) on the other hand, is absolutely vital if we are to proceed in an intellectually rigorous way, in such a way as to maximize the hope of progress. All this requires that we adopt and implement aim-oriented empiricism. No version of standard empiricism can do justice to the vital need to *revise* metaphysical blueprints (and associated criteria of explanatoriness and methodological principles of symmetry, invariance and conservation) in the light of improving empirical knowledge in the way specified explicitly by aim-oriented empiricism. In *suppressing* metaphysical assumptions involved in the scientific rejection of empirically successful aberrant theories, standard empiricism makes it impossible explicitly to *revise* such assumptions as science progresses. In attempting to make rational sense of science without (revisable) metaphysical presuppositions, traditional approaches to the problem of induction have sought to make rational what is actually profoundly *irrational*. Such attempts have sought to *justify the unjustifiable*! It is for this reason that the problem of induction has for so long remained unsolved.

Aim-oriented empiricism solves the problem of induction by providing a framework of metamethodological rules within which metaphysical blueprints and associated methodological principles or criteria of explanatoriness (required for the exclusion of empirically successful aberrant or *ad hoc* theories) are *revised* as science progresses. Science itself, in other words, pursued within the framework of aim-oriented empiricism, provides the means and the rationale for excluding empirically successful aberrant theories. (On this view, reasons for excluding *aberrant* theories—required for a solution to the problem of induction—are essentially the same as straightforward scientific reasons for excluding all *ad hoc* theories in science.) This solution creates, however, a new problem: how can it be justifiable to presuppose that the universe is comprehensible? Granted that the universe is comprehensible *in some way or other*, we can provide a rationale for assuming *physicalism*, the more or less specific way in which modern science presupposes that the world is comprehensible. For we can argue that the whole research programme of modern science, construed as presupposing physicalism, has been vastly more successful than any rival programme presupposing that the universe is comprehensible in some other way (animism or Aristotelianism, for example). But this argument itself presupposes that it is justifiable to assume that the universe is comprehensible *in some way or other*. Without this assumption, we have no grounds for dismissing research programmes with aberrant blueprints and aberrant versions of physicalism, which mimic the immense success of science. If the universe really is comprehensible in some way or other, however, then such aberrant versions of physicalism cannot be true. How can this essential assumption that the universe is comprehensible in some way or other be justified?

The answer, I suggest, is to construe this problem to be a part of the general problem of scepticism. The proper way to oppose scepticism is not to follow Descartes in trying to find some residue of our ostensible knowledge which is absolutely immune to doubt. There is no such residue: all our knowledge is fallible and conjectural, as Popper for one has tirelessly emphasized. In a sense, we should not seek to *oppose* scepticism at all; we should welcome it, as embodying the critical attitude which plays such a vital role in the *improvement* of knowledge, and which is, in particular, as Popper again has emphasized, such a vital ingredient of scientific method. Scepticism is only to be resisted when it threatens to destroy itself—when it leads to the undermining of the critical attitude and the very possibility of improving knowledge. We need, in this way, to be *sceptical* of scepticism, *critical* in our unleashing of criticism. We are rationally entitled to be sceptical of scepticism when it ceases to be fruitful and becomes sterile. What is vital is not the active doubting of everything but rather the doubting of what it is fruitful to doubt, in that this is capable of contributing to the improvement of knowledge. Whenever it can be shown that doubt cannot conceivably contribute to the improvement of knowledge, then it is rational to put doubt on one side. The thesis that the universe is comprehensible in some way or other is a case in point. It cannot be fruitful to doubt this thesis because the knowability of the universe—the possibility of improving knowledge—depends on this thesis being true. Nothing is to be gained from doubting the thesis, and everything is to be lost. We are rationally entitled to assume it is true. (For a more detailed development of this argument, see Maxwell [1984], Ch. 9 and (forthcoming), Ch. 4.)

At this point it may be objected that I have fallen into a gross error vividly depicted, as it happens, by van Fraassen ([1985], pp. 259–60). The error is to presuppose at the outset the success of science, show then that this requires some metaphysical thesis to be true, take this to justify the metaphysical thesis in question, and then, on that basis, justify science. 'From Gravesande's axiom of the uniformity of nature in 1717 to Russell's postulates of human knowledge in 1948, this has been a mug's game.' That is van Fraassen's assessment of this viciously circular argument. But is it not just this mug's game that I have been playing here?

It is not. In the first place, I accept without question that all our knowledge is conjectural in character. I do not attempt to justify science, physicalism or the comprehensibility thesis in any sense which converts conjectural knowledge into justified knowledge. I do however hold that it is rational to regard science as having enormously increased our (conjectural) knowledge about the world. I have argued at length that at any given stage in the development of science there is a more or less specific metaphysical thesis, asserting that the universe is comprehensible in such and such a way, *implicit* in what we take to be empirical and theoretical scientific knowledge. This neither presupposes, nor seeks to justify, the real success of science or the truth of the relevant metaphysical thesis. Indeed, far from seeking to *justify* the truth of the implicit metaphysical thesis, the current blueprint of science, I have done almost exactly the opposite: I have argued that the thesis is almost certainly *false*. The central message of aim-oriented empiricism is just that the current best metaphysical blueprint, implicit in basic concepts and principles of physics, is almost certainly *false*, and for that very reason needs to be made explicit so that it can be *criticized* and, we may hope, *improved*. We need to do this in order to satisfy elementary requirements for intellectual rigour, and in order to maximize the hope of scientific progress.

I have, it is true, argued that we ought to regard physicalism as being implicit in our contemporary scientific knowledge. But I have not presupposed the real success of science to justify the truth of physicalism; and I have not presupposed the truth of physicalism to justify science. Again, I have argued that we are rationally entitled to take the conjecture that the universe is comprehensible in some way or other to be a basic component of scientific knowledge. This argument does not presuppose the success of science; and nor does the argument, granted that it is successful, justify science. Granted that the universe is comprehensible in some way or other, it is not *this* in itself which justifies science but rather the fact that the research programme of modern science, presupposing physicalism, has met with far greater (apparent) empirical success than any rival research programme, presupposing that the universe is comprehensible in a way different from physicalism. Knowing that the universe is comprehensible in some way or other only provides a basis for rejecting grossly *aberrant* versions of modern science and associated grossly *aberrant* versions of physicalism which have all the empirical success of non-aberrant, actual science, if not more empirical success. (In unthinkingly *rejecting* such empirically successful aberrant possibilities we are in effect presupposing, on non-empirical grounds, that the world cannot be like that, that it cannot be grossly aberrant. All I have done is to make explicit this implicit presupposition.)

I conclude that in arguing that aim-oriented empiricism solves the problem of induction I have not been playing van Fraassen's mug's game. But even if I have, this does not affect adversely the main argument of part two of this paper. For strictly speaking all that I require for that argument is that aim-oriented empiricism is *necessary* for a solution to the problem of induction, not that it is *sufficient*. Indeed, all that I really require is that aim-oriented empiricism represents a more intellectually rigorous conception of science than standard empiricism in that it satisifes the elementary, general requirement for rigour of making explicit and so critizable what is substantial, influential and problematic, where standard empiricism fails to satisfy this requirement.[8]

Department of History and Philosophy of Science
University College London

[8] For a critical, but on the whole highly favourable assessment of aim-oriented empiricism, see Kneller [1978], pp. 80–95.

REFERENCES

Barker, S. F. [1957]: *Induction and Hypothesis*, Cornell University Press.

Churchland, P. M. and Hooker, C. A. (*eds.*) [1985]: *Images of Science*, University of Chicago Press.

Dieudonné, J. [1960]: *Foundations of Modern Analysis*, Academic Press.

Duhem, P. [1962]: *The Aim and Structure of Physical Theory*, Atheneum.

Glymour, C. [1980]: *Theory and Evidence*, Princeton University Press.

Goodman, N. [1954]: *Fact, Fiction and Forecast*, Athlone Press.

Harman, G. [1965]: 'The Inference to the Best Explanation', *Philosophical Review*, **74**, pp. 88–95.

Harman, G. [1968]: 'Enumerative Induction as Inference to the Best Explanation', *Journal of Philosophy*, **64**, pp. 529–33.

Howson, C. and Urbach, P. [1989]: *Scientific Reasoning: The Bayesian Approach*, Open Court.

Jeffreys, H. [1957]: *Scientific Inference*, Cambridge University Press.

Kneller, G. F. [1978]: *Science as a Human Endeavor*, Columbia University Press.

Kuhn, T. S. [1962]: *The Structure of Scientific Revolutions*, University of Chicago Press.

Lakatos, I. [1970]: 'Falsification and the Methodology of Scientific Research Programmes', in I. Lakatos and A. Musgrave (*eds.*), *Criticism and the Growth of Knowledge*, Cambridge University Press, pp. 91–195.

LAUDAN, L. [1990]: 'Normative Naturalism', *Philosophy of Science*, **48**, pp. 19–48.
LYCAN, W. [1988]: *Judgement and Justification*, Cambridge Univesity Press.
MAXWELL, N. [1966]: 'Physics and Common Sense', *British Journal for the Philosophy of Science*, **16**, pp. 295–311.
MAXWELL, N. [1968]: 'Understanding Sensations', *Australasian Journal of Philosophy*, **46**, pp. 127–46.
MAXWELL, N. [1972]: 'A Critique of Popper's Views on Scientific Method', *Philosophy of Science*, **39**, pp. 131–52.
MAXWELL, N. [1974]: 'The Rationality of Scientific Discovery', Parts I and II, *Philosophy of Science*, **41**, pp. 123–53 and 247–95.
MAXWELL, N. [1976]: *What's Wrong With Science?*, Bran's Head Books.
MAXWELL, N. [1977]: 'Articulating the Aims of Science', *Nature*, **265**, p. 2.
MAXWELL, N. [1979]: 'Induction, Simplicity and Scientific Progress', *Scientia*, **114**, pp. 629–53.
MAXWELL, N. [1980]: 'Science, Reason, Knowledge and Wisdom: A Critique of Specialism' *Inquiry*, **23**, pp. 19–81.
MAXWELL, N. [1984]: *From Knowledge to Wisdom: A Revolution in the Aims and Methods of Science*, Basil Blackwell.
MAXWELL, N. [1986]: 'The Fate of the Enlightenment: Reply to Kekes', *Inquiry*, **29**, pp. 78–82.
MAXWELL, N. [1991]: 'How Can we Build a Better World?', in J. Mittelstrass (*ed.*), *Einheit der Wissenschaften*. Walter de Gruyter, pp. 388–427.
MAXWELL, N. (forthcoming): *The Rationality of Scientific Discovery*, Cambridge University Press.
NORTH, J. D. [1965]: *The Measure of the Universe*, Clarendon Press.
POPPER, K. R. [1959]: *The Logic of Scientific Discovery*, Hutchinson.
POPPER, K. R. [1963]: *Conjectures and Refutations*, Routledge & Kegan Paul.
STOVE, D. [1982]: *Popper and After*, Pergamon Press.
VAN FRAASSEN, B. C. [1980]: *The Scientific Image*, Clarendon Press.
VAN FRAASSEN, B. C. [1985]: 'Empiricism in the Philosophy of Science', in Churchland and Hooker (*eds.*) [1985].
WATKINS, J. [1984]: *Science and Scepticism*, Princeton University Press.